Atty. Dkt. No.: 035905/0103

# U.S. PATENT APPLICATION

for

## Nonvolatile Memory on SOI and Compound Semiconductor Substrates and Method of Fabrication


Inventors:

Andrew J. Walker, Mark G. Johnson, N. Johan Knall,

Igor G. Kouznetsov and Christopher J. Petti



Prepared by:
Foley & Lardner
3000 K St. N.W.
Washington D.C. 20007
(202) 672-5300

# Nonvolatile Memory on SOI and Compound Semiconductor Substrates and Method of Fabrication

[0001] This application is a continuation-in-part of U.S. Application Serial Number 09/814,727, filed on March 21, 2001, which is a continuation of U.S. Application Serial Number 09/560,626, filed on April 28, 2000, both of which are incorporated by reference in their entirety.

## FIELD OF THE INVENTION

[0002] The present invention is directed generally to semiconductor devices and methods of fabrication and more particularly to a nonvolatile memory array and method of fabrication.

## BACKGROUND OF THE INVENTION

[0003] Field programmable nonvolatile memory arrays such as programmable read only memories (PROMs), electrically programmable read only memories (EPROMs), and electrically erasable programmable read only memories (EEPROMs) require driver circuits (also known as peripheral circuits) which select a particular device in the array to write or read data from the particular device. Frequently, such driver circuits are formed in a bulk monocrystalline silicon substrate, while the memory arrays are formed above the driver circuits. Such a memory array is disclosed in U.S. patents 6,034,882 and 6,185,122, incorporated herein by reference. However, forming the driver circuits in the bulk monocrystalline silicon substrate may be undesirable for some applications of the memory array.

## BRIEF SUMMARY OF THE INVENTION

[0004] A preferred embodiment of the present invention provides a nonvolatile memory array, comprising an array of nonvolatile memory devices, at least one driver circuit, and a substrate, wherein the at least one driver circuit is not located in a bulk monocrystalline silicon substrate.

[0005] Another preferred embodiment of the present invention provides a nonvolatile memory array, comprising a monocrystalline silicon substrate, at least one driver circuit formed above the substrate, and an array of nonvolatile memory devices formed above the substrate.

[0006] Another preferred embodiment of the present invention provides a method of making a nonvolatile memory array, comprising forming at least one driver circuit above a substrate or in a semiconductor substrate other than a monocrystalline silicon substrate, and forming an array of nonvolatile memory devices.



## BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Figures 1-3 are schematic side cross sectional views of a memory array according to the first and second preferred embodiments of the invention.

[0008] Figure 4A is a three dimensional view of a rail stack antifuse memory array according to the third preferred embodiment of the invention.

[0009] Figure 4B is a side cross sectional view of a rail stack antifuse memory array according to an alternative aspect of the third preferred embodiment of the invention.

[0010] Figure 5 is a three dimensional view of a rail stack EEPROM memory array according to the fourth preferred embodiment of the invention.

[0011] Figures 6 and 7 are three dimensional views of two terminal pillar devices according to the fifth preferred embodiment of the invention.

[0012] Figure 8 is a three dimensional view of an EEPROM pillar device according to the sixth preferred embodiment of the invention.

[0013] Figure 9 is a three dimensional view of an TFT EEPROM array according to the seventh preferred embodiment of the invention.

[0014] Figures 10A is side cross sectional view of an FN tunneling flash memory array according to the eighth preferred embodiment of the invention. Figure 10B is a side cross sectional view along line B-B in Figure 10A.

[0015] Figure 11 is a side cross section views of a TFT CMOS EEPROM array according to the ninth preferred embodiment of the invention.

[0016] Figures 12 to 14 are circuit diagrams of logic circuits using the TFT CMOS EEPROM array of the ninth preferred embodiment.

[0017] Figure 15 is a schematic side cross sectional view of a memory array monolithically integrated with a non-memory device.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0018] The present inventors have realized that the driver circuit of the nonvolatile memory array does not have to be located in a bulk monocrystalline silicon substrate. Instead, the driver circuit may be

formed in a silicon on insulator (SOI) substrate or in a compound semiconductor substrate, depending on the type of electronic device in which the array is to be incorporated.

**[0019]** Figure 1 illustrates a schematic overview of a nonvolatile memory array 1 according to the first preferred embodiment of the present invention. In Figure 1 at least one driver circuit 2 is formed in an SOI substrate, which contains an insulating surface 3. An array of nonvolatile memory devices 4 is located above the at least one driver circuit 2. Alternatively, the array of nonvolatile memory devices 4 may be located adjacent to or below the driver circuit 2 in the SOI substrate.

**[0020]** The preferred SOI substrates of the first preferred embodiment are schematically illustrated in Figures 2A and 2B. The SOI substrate includes a silicon layer 5 over the insulating surface 3. The silicon layer 5 may be a single crystal silicon layer, a polycrystalline silicon (i.e., "polysilicon") layer or an amorphous silicon layer. The at least one driver circuit 2 is located in the silicon layer 5. If there are a plurality of driver circuits 2 in the silicon layer 5, then the driver circuits may be isolated from each other by field oxides, trench isolation structures and/or any other known isolation structures. The insulating surface 3 below the silicon layer 5 may be an insulating layer 6 formed on or over a semiconductor substrate 7 or an insulating substrate 8, as shown in Figures 2A and 2B, respectively.

**[0021]** In a first preferred aspect of the first embodiment, the SOI substrate includes a silicon layer 5 formed on an insulating layer 6, such as silicon oxide, silicon nitride, silicon oxynitride and/or other insulting layers. The insulating layer 6 is located on or over a bulk monocrystalline silicon substrate 7, as shown in Figure 2A. Thus, the at least one driver

circuit 2 is located in layer 5 above the bulk monocrystalline silicon substrate 7.

**[0022]** The SOI substrate (i.e., elements 5, 6, and 7) shown in Figure 2A may be formed by the SIMOX (separation by implantation of oxygen) method. The SIMOX method includes providing a monocrystalline silicon substrate 7 and implanting oxygen below the surface of the substrate 7. The substrate 7 is then annealed to form a silicon oxide layer 6 in the substrate. Since the oxygen was implanted below the substrate 7 surface, a single crystal silicon layer 5 (which used to comprise the top section of the substrate 7) remains above the silicon oxide layer 6. The at least one driver circuit 2 is then formed in the single crystal silicon layer 5. The at least one driver circuit 2 preferably contains MOS (metal oxide semiconductor) and CMOS (complementary MOS) transistors and is preferably fabricated in substrate using ordinary MOS and CMOS fabrication techniques.

**[0023]** Alternatively, the SOI substrate may be formed by a seeded lateral epitaxy method. In this method, a monocrystalline silicon substrate 7 is provided. An insulating layer 6, such as a silicon oxide layer, is formed over the substrate 7. One or more windows or vias to the substrate 7 are formed in layer 6. A silicon layer 5 is deposited over the silicon oxide layer 6, such that it contacts the substrate 7 through the windows in layer 6. The silicon layer 5 may be deposited as a single crystal silicon layer over the silicon oxide layer using the substrate as a seed. Alternatively, the silicon layer 5 may be deposited as an amorphous or polycrystalline silicon layer, and then recrystallized by laser or thermal annealing into a single crystal layer using the substrate 7 as a seed. Then, the at least one driver circuit 2 is formed in the single crystal silicon layer 5.

[0024] Alternatively, the silicon layer 5 may comprise a polycrystalline silicon or an amorphous silicon layer formed over the insulating layer 6. In this case, the insulating layer 6, such as a silicon oxide, silicon nitride or silicon oxynitride interlayer insulating layer is formed over the substrate 7. The polycrystalline or amorphous silicon layer 5 is then deposited over the insulating layer 6, and the driver circuit 2 is formed in the layer 5. If desired, the crystallinity of the layer 5 may be improved by laser and/or thermal annealing. Thus, for example, an amorphous silicon layer may be converted into a polycrystalline silicon layer. If desired, a crystallization catalyst material, such as Ge, or a transition metal, such as Ni, Pt, Pd, etc., or their silicides, may be used as a seed for the crystallization of the amorphous silicon layer.

[0025] In a second preferred aspect of the first embodiment, the SOI substrate includes a silicon layer 5 formed on or over an insulating substrate 8, as shown in Figure 2B. The insulating substrate 8 may be a glass, plastic or ceramic substrate. Glass substrates include, for example, silicate glass, aluminosilicate glass, boroaluminosilicate glass, germanate glass and various glass-ceramics (i.e., glasses which contain at least 5% crystal phase). Plastic substrates are preferably flexible or bendable, and include, for example, polyimide, PTFE and various other polymer based materials. Ceramic substrates include, for example, sapphire, alumina, zirconia, yttria stabilized zirconia and quartz (crystalline silica), etc. It should be noted that it is possible to form a single crystal silicon layer directly on a sapphire substrate.

[0026] The silicon layer 5 formed on the insulating substrate 8 may be a single crystal silicon layer formed by the wafer bonding method. This method includes providing a temporary monocrystalline silicon substrate (not shown) and forming the at least one driver circuit 2 in this substrate. The temporary substrate is then selectively removed from

below the driver circuit 2, such that only the single crystal silicon layer 5 in which the driver circuit 2 is located remains. The temporary substrate may be removed by selective wet etching which preferentially etches the first conductivity type substrate to the second conductivity type layer 5 in which the driver circuit 2 is located. Alternatively, the temporary substrate may be removed by non-selective etch back or chemical mechanical polishing which is stopped before any portion of the driver circuit is removed, by using an etch stop layer and/or a timed etchback or polishing. After the temporary substrate is removed, a permanent insulating substrate 8 (such as a glass substrate) is attached to the at least one driver circuit 2.

**[0027]** Alternatively, the silicon layer 5 may be a polycrystalline or an amorphous silicon layer formed directly over an insulating substrate 8. If desired, an optional impurity blocking insulating layer, such as a silicon oxide, silicon nitride or aluminum oxide layer may be formed between the insulating substrate 8 and the silicon layer 5. If desired, the crystallinity of the layer 5 may be improved by laser and/or thermal annealing. Thus, for example, an amorphous silicon layer may be converted into a polycrystalline silicon layer. If desired, a crystallization catalyst material, such as Ni, Ge, Pt, Pd, etc., may be used as a seed for the crystallization of the amorphous silicon layer.

**[0028]** In an alternative aspect of the first preferred embodiment, at least a part, and preferably all of the at least one driver circuit 2 is formed within the memory array or above the memory array. Thus, the driver circuit 2 is also formed in an SOI substrate, which comprises an amorphous or polycrystalline semiconductor (i.e., silicon) layer(s) which is separated from a monocrystalline semiconductor or insulating substrate by one or more interlayer insulating layers and/or one or more device levels of the memory array. In this aspect of the first embodiment, the at

least one driver circuit is vertically integrated with the memory array to increase the device density.

**[0029]** In a second preferred embodiment of the present invention, the at least one driver circuit 2 is formed in a III-V, II-VI or IV-IV semiconductor substrate 9, as shown in Figure 3. The array of nonvolatile memory devices (not shown in Figure 3) is formed above or adjacent to the at least one driver circuit 2. For example, the III-V semiconductor substrate may be a GaAs, InP or GaN substrate. The II-VI semiconductor substrate may be a ZnSe, CdSe or CdS substrate. The IV-IV substrate may be a silicon carbide ("SiC") or a SiGe substrate.

**[0030]** In the arrays of the first and second embodiments, the at least one driver circuit 2 may include a decoding circuit, a sensing circuit, a programming circuit and/or other logic circuits. The array of nonvolatile memory devices 4 comprises an array of PROMs, EPROMs or EEPROMs. The array 4 may be a two or a three dimensional array. Preferably, the array of nonvolatile memory devices comprises a monolithic three dimensional array of memory devices. The term "monolithic" means that layers of each level of the array were directly deposited on the layers of each underlying level of the array. Thus, a first interlayer insulating layer is formed over the at least one driver circuit, at least one first semiconductor layer is deposited over the first interlayer insulating layer, and a first array of PROMs, EPROMs or EEPROMs is formed in the at least one first semiconductor layer. Then, a second interlayer insulating layer is formed over the first array of PROMs, EPROMs or EEPROMs. At least one second semiconductor layer is formed over the second interlayer insulating layer. A second array of PROMs, EPROMs or EEPROMs is formed in the at least one second semiconductor layer to form a monolithic three dimensional array. Additional array levels may be formed in the same fashion if desired. In contrast, two dimensional arrays may

be formed separately and then packaged together to form a non-monolithic memory device.

[0031] The various nonvolatile memory devices which may be formed in the array of memory devices 4 are described below with respect to the third through seventh preferred embodiments. It should be noted that memory devices other than those described below may formed in the array instead. For example, the memory devices may comprise the devices disclosed in U.S. Patents 5,825,046, 6,075,719, 6,087,674 and 6,141,241, incorporated herein by references.

[0032] In a third preferred embodiment of the present invention, the array of nonvolatile memory devices 4 comprises a three dimensional array of antifuses. The array of antifuses preferably comprises a first set of rail stack conductors, a second set of rail stack conductors extending in a different direction than the first set of rail stack conductors, and an insulating layer disposed between the first and the second sets of rail stacks, as illustrated in Figure 4.

[0033] The array of Figure 4A may have any number of levels of memory cells, such as 2 to 8 levels. Each level includes a first plurality of parallel spaced-apart rail-stacks running in a first direction and a second plurality of rail-stacks or conductors running in a second direction. Preferably, the first rail-stacks run perpendicular to the second conductors/rail-stacks and hence form a right angle at their intersections.

[0034] A bit can be stored at each of the intersections of rail-stacks. However, there are no physically discrete individual memory cells at the intersections, rather memory cells are defined by the rail-stacks and intermediate layers. This makes it easier to fabricate the memory array. The term "memory cell" is intended broadly to encompass physically discrete elements or elements that are defined by rail-stacks and

intermediate layers, or any other localized region where a bit can be stored. When the array is fabricated all the bits are in the zero (or one) state and after programming, the programmed bits are in the one (or zero) state.

**[0035]** In the embodiment of Figure 4A, several rail-stacks are illustrated in the partial cross-section of the array. For instance, a rail-stack 16 is shown at one height and a half rail-stack 18 is shown at a second height above the first height. Also, half rail-stacks are disposed between rail-stack 16 and a substrate 10.

**[0036]** These lower rail-stacks run in the same direction as the half rail-stack 18. A bit is stored at the intersection of rail-stacks and, for instance, a "cell" is present between the rail-stacks as shown within the bracket 17 and another within the bracket 19. Each of these brackets spans a memory level.

**[0037]** The substrate 10 may comprise an SOI substrate of the first embodiment or a compound semiconductor substrate of the second embodiment. The at least one driver circuit 2 is fabricated in substrate 10 under the memory array 4 using, for instance, ordinary MOS and CMOS fabrication techniques. Vias are used to connect conductors within the rail-stacks to the substrate to allow access to each rail-stack in order to program data into the array and to read data from the array. For instance, the circuitry within the substrate 10 may select the rail-stack 16 and the rail stack 18 to program or to read a bit associated with the intersection of these rail-stacks. Alternatively, the at least one driver circuit 2 may be formed within or above the memory array, if desired.

**[0038]** As shown in Figure 4A, an insulating layer 12 is formed over the substrate 10 containing the driver circuit(s) in order that the array may be fabricated above the substrate. This layer may be planarized

with, for instance, chemical-mechanical polishing (CMP) to provide a flat surface upon which the array may be fabricated.

**[0039]** Following this, a conductive layer 14 is formed over the substrate. As will be seen, conductive layers are used within the rail-stacks and these layers and the resultant conductors may be fabricated from elemental metals such as tungsten, tantalum, aluminum, copper or metal alloys such as MoW. Metal silicides may also be used such as $TiSi_2$, $CoSi_2$ or a conductive compound such as TiN, WC may be used. A highly doped semiconductor layer such as silicon is also suitable. Multiple layer structures may be used selecting one or more of the above.

**[0040]** Following the deposition of a conductive layer, a layer of semiconductor material 15, such as silicon, is formed over the conductive layer. This is typically a polysilicon layer; however, an amorphous layer may be used. Other semiconductor materials may be used such as Ge, GaAs, etc. In the embodiment of Figure 4A, this semiconductor layer is highly doped and, as will be seen, forms one-half a diode. After masking and etching steps, half rail-stacks are formed. These rail-stacks are "half" or partial rail-stacks since they are approximately half the thickness of the rail-stacks used in the next level.

**[0041]** Following this, in the embodiment of Figure 4A, a material for the antifuses used to program the array is deposited. In one embodiment, the layer 20 is a dielectric such as silicon dioxide which is deposited by chemical vapor deposition (CVD) in a blanket deposition over the half rail-stacks and filling the space between the rail-stacks to form a dielectric fill. Preferably, the fill insulating material, such as silicon oxide, is blanket deposited to fill in the spaces between the rail stacks which include layers 14 and 15. The fill material is then planarized to expose the top surface of the semiconductor layer 15 in the rail stacks, and the

antifuse layer 20, such as silicon oxide, is deposited over the rail stacks and the fill material. In another alternative aspect of this embodiment, the layer 20 is selectively grown on the upper surface of the silicon layer 15 and only exists on the rail-stacks.

**[0042]** Now a full set of memory array rail-stacks is formed on the layer 20. This comprises first the deposition of a lightly doped silicon layer 21 doped with a conductivity-type dopant opposite to that used for the silicon layer 15, a heavily doped silicon layer 22 doped also opposite to the layer 15, a conductive layer 23 and a heavily doped silicon layer 24 doped with the same conductivity-type dopant as layers 21 and 22. After masking and etching, the rail-stacks shown in Figure 4A, such as rail-stack 16, are formed. These rail-stacks are oriented, as illustrated, in a direction perpendicular to the rail-stacks above and below them.

**[0043]** While not shown in Figure 4A, the spaces between the rail-stacks are filled with a dielectric such as silicon dioxide. Then the rail-stacks and fill are planarized by chemical mechanical polishing (CMP). In another embodiment spin-on-glass (SOG) is used to fill the voids. In this case chemical planarization (i.e., etch back) can be used. Other fill and planarization methods can be used.

**[0044]** After formation of the rail-stacks another antifuse layer 26 is formed, for instance from a dielectric such as silicon dioxide, silicon nitride, silicon oxynitride, amorphous carbon or other insulating materials or combinations of materials. Also an updoped layer of silicon may be used for the antifuse layer.

**[0045]** Now another layer of rail-stacks is defined and only half rail-stacks are shown in Figure 4A at this upper level. This half rail-stack comprises a silicon layer 28 doped with a conductivity-type dopant opposite to that of layer 24. This is a lightly doped layer. Another silicon

layer 30 is formed on layer 28 and this layer is doped with the same conductivity-type dopant as layer 28; however, it is more heavily doped. Then a conductive layer 31 is formed above the layer 30.

**[0046]** Half rail-stacks are used at the very upper-most level of the array and at the very lowest level of the array. In between the half rail-stacks, full rail-stacks, such as rail-stack 16, are used throughout the array.

**[0047]** It should be noted that the silicon layers disposed on the conductive layers extend the entire length of the rail-stacks in the embodiment of Figure 4A and are uninterrupted, except possibly where vias are used to provide a conductive path to the substrate 10.

**[0048]** In Figure 4A a path 32 is illustrated from a lower conductor in level 17 to an upper conductor in this level found in the rail-stack 18. This path is accessed in one embodiment through decoding circuitry in the substrate for both programming and reading of data into and from the array for one bit.

**[0049]** For instance, to program the bit, a relatively high voltage, e.g. 5-20V, is applied between the conductors to forward-bias the diode between these conductors. This relatively high voltage causes a breach in the layer 26 creating a diode. Without this high voltage, the layer 26 remains an insulator. Thus, by selecting pairs of conductors, diodes can be selectively formed so as to program the array. While programming the array with the layers adjacent to the antifuse material being forward-biased is currently preferred, it is also possible to program using a reverse-biasing potential.

**[0050]** To sense the data programmed into the array, a voltage lower than the programming voltage is used. This lower voltage is

applied so as to forward-bias the diode of the cell being accessed and thus allow a sense amplifier to determine whether or not the layer 26 is intact between the rail-stacks. Note that "sneak" or parasitic paths in the array which would interfere with the sensing will include a reverse-biased diode.

**[0051]** Also, the anode and cathode of the diodes are reversed at each of the successive antifuse layers. This facilitates programming and sensing, since all of the conductors at each level are either bitlines or wordlines. And, for instance, conductors at one height will serve as bitlines for two levels and conductors at the next height serve as wordlines for two levels. This simplifies the decoding and sensing and more importantly reduces processing.

**[0052]** Some diode types may exhibit a higher leakage current than other diode types. Consequently, it may be desirable to form an array with only a single diode type. Specifically, the P- / N+ diodes have a higher leakage current than P+ / N- diodes. Figure 4B illustrates an array of a preferred aspect of the third embodiment where, if the antifuse layer is breached, all the diodes will have a P+ / N- junction. Thus, there will be no diodes with a P- / N+ junction.

**[0053]** In Figure 4B, three rail-stacks 120, 121, and 122 are illustrated which will create a P+ / N- diodes when an antifuse layer is breached. The first rail-stack 120 comprises a first P+ semiconductor (i.e., silicon) layer 125, a conductor 126, a second P+ layer 127 and an antifuse layer 128. The thickness of layers 125-127 may be 400 to 3000 Å, such as 1,000 / 500 / 1,000Å, for layers 125, 126 and 127 respectively. The antifuse layer 128 may be about 20-40 Å, preferably about 30Å thick.

[0054] The second rail-stack 121 comprises a first N- semiconductor (i.e., silicon) layer 129, a first N+ semiconductor layer 130, a conductor 131, a second N+ layer 132 and a second N- layer 133. The thickness of layers 129-133 may be 400 to 3000 Å, such as 2,000 / 500 / 500 / 500 / 2000 Å for layers 129, 130, 131, 132 and 133 respectively.

[0055] The third rail-stack 122 contains the same layers as the first rail-stack 120 in reverse order. The third rail stack includes a first P+ semiconductor (i.e., silicon) layer 139, a conductor 140, a second P+ layer 142 and an antifuse layer 143.

[0056] As discussed above, the semiconductor layers may comprise polysilicon or amorphous silicon. The conductors may be heavily doped polysilicon, metal, silicide or combinations thereof. A CMP planarized dielectric fill 144, such as silicon dioxide, is provided in the spaces between the rail-stacks.

[0057] As can be seen from Figure 4B, if the antifuse layer 128 is breached, the diodes between the conductors 126 and 131 are all P+ / N- type. Similarly, the diodes in the next level between the conductors 131 and 140 are again all P+ / N- type. The rail-stacks shown are used throughout the memory array so that the entire array has only P+ / N- type diodes in its memory cells.

[0058] The diodes in the illustrated rail-stacks of Figure 4B are forward biased towards the conductor 131 and the conductor 141. If need be for a particular application, the diodes can be oriented identically, that is, with all their anodes (or cathodes) pointing upwardly. This can be obtained for the P+ / N- type diodes by having both a P+ doped and N- doped semiconductor layer in each of the rail-stacks. For instance, layers 132 and 133 would be replaced with a P+ layer and layer 142 would be

replaced with N- and N+ layers. This still maintains only one type of diode (P+ / N-) throughout the array.

**[0059]** While Figure 4B shows that after the antifuse layer is breached, only P+ / N- diodes will be created, an array with only P- / N+ type diodes can be fabricated by replacing the P+ layers with N+ layers and replacing the N+ and N- layers with P+ and P- layers, respectively. Also, the array can have the anodes (or cathodes) vertically aligned as discussed above for the P+ / N- type diodes.

**[0060]** According to the fourth preferred embodiment of the present invention, the array of non-volatile memory devices comprises a three dimensional array of rail stack EEPROMs. These EEPROMs are three terminal devices in contrast to the two terminal antifuse devices of the third embodiment. The array comprises a first plurality of spaced-apart conductors disposed at a first height above the substrate in a first direction. The array also comprises a second plurality of spaced-apart rail-stacks disposed above the first height in a second direction different from the first direction, each rail-stack including a semiconductor film of a first conductivity type in contact with said first plurality of spaced-apart conductors, a local charge storage film disposed above the semiconductor film and a conductive film disposed above the local charge storage film.

**[0061]** Figure 5 illustrates a three dimensional array of rail stack EEPROMs according to the fourth preferred embodiment of the present invention. The array 40 includes a first plurality of spaced-apart conductors such as n+ doped polysilicon bit lines 42, 44, 46, 48 disposed in a first direction above the substrate (not shown). A second plurality of spaced-apart "rail stacks" 50, 52 are disposed in a second direction different from the first direction (and preferably orthogonally) at a second height above the substrate so that they are above bit lines 42,

44, 46 and 48 and in contact therewith at intersection points 54, 56, 58, 60, 62, 64, 66, 68. Each rail stack 50, 52 in this embodiment includes at least a layer of p- doped polysilicon 70. Over layer 70 is disposed a charge storage medium 72, such as a dielectric isolated floating gate, an ONO dielectric film (i.e., a $SiO_2/Si_3N_{4-x}O_{1.5x}/SiO_2$ film, where $0 \leq x \leq 1$) or an insulating layer containing conductive nanocrystals. A conductive wordline 74 which may comprise n+ doped polysilicon is disposed over the charge storage medium 72. A planarized oxide material (not shown) may be deposited in the spaces between adjacent bit lines and rail stacks. A conventional chemical mechanical polishing (CMP) process may be used to accomplish this. If desired, the bit lines 42, 44, 46, 48 may comprise p+ doped rather than n+ doped polysilicon, while layer 70 may comprise n- doped rather than p- doped polysilicon.

**[0062]** The memory array structure of Figure 5 can be easily extrapolated to three dimensions. To do this, an interlayer insulating layer is placed over the rail stacks 50, 52 after the CMP process. This layer prevents shorting one set of wordlines with the next set of bit lines. Then another layer of bit lines 42, 44, 46, 48 is constructed over the interlayer insulating layer followed by an oxide deposition and a CMP step, followed by a deposition of another set of rail stacks. This process can be repeated a number of times, as desired, to form two or more device levels (such as eight levels, for example).

**[0063]** According to the fifth preferred embodiment of the present invention, the array of non-volatile memory devices 4 comprises a three dimensional array of two terminal pillar memory devices. Referring to Figure 6, one device of the three-dimensional memory is illustrated. The device includes a conductor 81 at one level and a conductor 80 at the next level in the array. A pillar structure is formed in alignment with the conductors 80 and 81. This pillar structure forms a cell in accordance

with the present embodiment. Specifically, referring to Figure 6, the cell includes a steering element and a storage stack. The steering element comprises a junction diode which contains a p+ region 82 and an n- region 10. The storage stack comprises a tunnel oxide region 84, a charge trapping region 86 and a blocking oxide 85. The charge trapping region 86 may comprise silicon nitride, silicon oxynitride, an ONO dielectric film or silicon nanocrystals in an insulating layer. The conductors 80 and 81 are shared with cells disposed above and below the single cell shown in Figure 6.

**[0064]** Figure 7 shows another preferred aspect of this embodiment where again there are spaced-apart, parallel conductors at one level, such as conductor 91, and parallel, spaced-apart conductors at the next level, such as conductor 90. A pillar structure is again fabricated between the conductors 90 and 91. The difference, however, between the structure of Figure 6 and Figure 7, is that the storage stack comprising the blocking oxide 93, charge trapping region 94 and tunnel oxide 95 is disposed between the p and n regions of the diode. Specifically, the p+ region 92 of the diode is in contact with the blocking oxide 93 and the n- region 96 is in contact with the tunnel oxide 95.

**[0065]** According to the sixth preferred embodiment of the present invention, the array of non-volatile memory devices 4 comprises a three dimensional array of three terminal pillar memory devices, such as pillar EEPROMs. These devices differ from the devices of the previous embodiment in that they contain three rather than two terminals. A three terminal nonvolatile stackable pillar memory device 100 is illustrated in Figure 8. Pillar memory device 100 includes a first contact region 102 formed on a first input/output (I/O) 103 conductor (i.e., electrode), which is formed above a plane (x-y) of a substrate 101. A semiconductor body 104 is formed directly on the first contact region 102 and a second

contact region 106 is formed directly on the body 104. Preferably, the contact regions 102 and 104 are heavily doped semiconductor regions of a first conductivity type (i.e., source and drain regions), while the body is a lightly doped semiconductor region of a second conductivity type (i.e., a channel). The semiconductor regions 102, 104 and 106 preferably comprise doped polysilicon. A second I/O conductor 116 (i.e., electrode) is formed on the second contact region 106. The first contact region 102, the body 104, and the second contact region 106 are each vertically aligned with one another to form a pillar 108.

**[0066]** Adjacent to and in contact with body 104 is a charge storage medium or region 110. A control gate 112 is formed adjacent to and in direct contact with the charge storage medium or region 110. The control gate 112 and charge storage medium 110 are constructed so that they lie laterally adjacent to pillar 108 so that they may electrically communicate with the pillar 108. The charge storage medium is the region that electrically screens the control gate and the channel region addressed by the control gate. The charge storage medium may comprise a silicon oxide/nitride/oxide ("ONO") dielectric film, conductive nanocrystals in an insulating layer or a floating gate located between a tunnel dielectric layer and a control gate dielectric layer.

**[0067]** During read operations of device 100, when a conductive channel is formed in body 104, current 114 flows vertically (z) (or perpendicular) with respect to the plane (x-y) of the substrate 101 above which pillar memory device is formed. By creating a memory device with a "vertical" read current path, the pillar memory cell of the present embodiment can be easily stacked in a three dimensional array with source/drain conductors 103 and 116 running parallel or perpendicular to each other and parallel to the plane of the substrate 101 without requiring the use of vertical interconnect strategies for the source and drain

connections. The conductor 112 to the control gate may be run vertically (as shown in Figure 8) or horizontally.

**[0068]** Although memory device 100 shown in Figure 8 includes a charge storage medium 110 and a control gate 112 formed on only one side or surface of pillar 108, it is to be appreciated that the pillar memory device of the present embodiment can be fabricated so that the entire body of the pillar 108 is surrounded by a single charge storage member 110 and a single control gate 112. Additionally, each surface of the pillar 108 can have an independently controlled charge storage member and control gate and thereby enable multiple bits of data to be stored in a single pillar memory device. The use of multiple charge storage members and control gates enables the storage of multiple values on a single pillar device by determining how much of the channel is exposed to charge. Additionally, each face of body 104 of pillar 108 can have different doping densities to create different threshold voltages for each face to further enable the pillar memory to store additional states and therefore additional bits.

**[0069]** According to the seventh preferred embodiment of the present invention, the array of non-volatile memory devices 4 comprises a three dimensional array of thin film transistor ("TFT") EEPROMs. This array comprises a plurality of vertically separated device levels, each level comprising an array of TFT EEPROMs. Each TFT EEPROM includes a channel, source and drain regions, a control gate, and a charge storage region between the channel and the control gate. The array also comprises a plurality of bit line columns in each device level, each bit line contacting the source or the drain regions of the TFT EEPROMs. The array further comprises a plurality of word line rows in each device level, and at least one interlayer insulating layer located between the device levels.

**[0070]** Figure 9 illustrates a preferred three dimensional memory array 200 according to the seventh preferred embodiment of the present invention. The three dimensional memory array 200 is a three dimensional array of TFT EEPROMs. Each TFT EEPROM contains source and drain regions 217, a channel 219, a control gate 243, control gate sidewall spacers (not shown for clarity in Figure 9) and a charge storage region 207 between the channel and the control gate. The channel 219 of each TFT EEPROM comprises amorphous silicon or polysilicon. The charge storage region 207 may comprise a silicon oxide/nitride/oxide ("ONO") dielectric film, isolated nanocrystals or a floating gate located between a tunneling dielectric and a control gate dielectric layer.

[0071] The memory array 200 also contains a plurality of bit line columns 225. Each bit line 225 contacts the source or the drain regions 217 of a plurality of TFT EEPROMs. The columns of the bit lines 225 extend substantially perpendicular to the source-channel-drain direction of the TFT EEPROMs (i.e., a small deviation from the perpendicular direction is included in the term "substantially perpendicular"). The bit lines 225 in each device level are shaped as rails which extend under the intergate insulating layer 227. The bit lines include the buried diffusion regions formed during the source and drain 217 doping steps and the overlying silicide 223 layers. The source and drain regions 217 are formed in the bit lines 225 where the word lines 241 intersect (i.e., overly) the bit lines 225. The doped source and drain regions 217 are located adjacent to the TFT EEPROM channel regions 219.

[0072] The memory array 200 also includes a plurality of word line rows 241. Each word line 241 contacts the control gates 243 of a plurality of TFT EEPROMs 200. Alternatively, the word lines can themselves comprise the control gates. The rows of word lines extend substantially parallel to the source-channel-drain direction of the TFT

EEPROMs. The plurality of word lines 241 are self aligned to the control gates 243 of the array of TFT EEPROMs or the word lines themselves comprise the control gates. If floating gates, but not control gates are included in the array, then the word lines are self aligned to the floating gates and to the control gate dielectric. The word lines are also self aligned to the channel and the charge storage regions of the TFT EEPROMs located below the respective word lines.

[0073] Each device level 245 of the array is separated and decoupled in the vertical direction by an interlayer insulating layer 203. The interlayer insulating layer 203 also isolates adjacent word lines 241 and adjacent portions of the channels 219 below the respective word lines 241 in each device level 245. The effective cell area per bit in the resulting three dimensional memory array is about $2f^2/N$, where N is number of device levels (i.e., $N=1$ for a two dimensional array and $N>1$ for a three dimensional array).

[0074] Each level of the TFT EEPROMs may be formed by forming a plurality of gate electrodes 209 on the active silicon areas. The source and drain regions 217 are implanted into the active silicon areas using the gates 209 as a mask. The remaining silicon active areas form the TFT channel regions 219. Sidewall spacers (not shown) are formed on the gates 209. Then, silicide regions 223 are formed on the source and drain regions 217 by the salicide method. The intergate insulating layer 227 is deposited over the gates 209 and planarized to expose the gates 209. If desired, a sacrificial blocking layer may also be formed over the gates 209. The blocking layer is removed after layer 227 is planarized to expose the gates 209. Then a conductive material is deposited and patterned to form the word lines 241 and control gates 243. The channels 219 and the charge storage region 207 are also patterned during the same etching step using the same photoresist mask.

[0075] The memory devices of the preferred embodiments of the present invention may be arranged in a three dimensional virtual ground array (VGA) nonvolatile flash memory. The devices may also be formed in nonvolatile flash memory architectures other than VGA, such as NOR-type memory and Dual String NOR (DuSNOR) memory architectures.

[0076] In a VGA illustrated in the previous embodiments, the programming of each EEPROM occurs by hot carrier injection. In hot carrier injection, a voltage is placed across a diode (i.e., between a source and a drain of a TFT EEPROM). The hot carriers (i.e., hot electrons and holes) that are travelling from source to drain through the channel of the TFT EEPROM are injected into the charge storage region which is disposed adjacent to the channel. This procedure is a relatively high power event.

[0077] For low power portable applications where both program/erase and read power are important, a flash nonvolatile memory using Fowler-Nordheim tunneling ("FN tunneling") for both program and erase may be used. FN tunneling results from applying a voltage across a dielectric. Thus, in a TFT EEPROM, a voltage is applied between a control gate and a source and/or a drain region of the TFT, for writing and erasing the TFT EEPROM. This is in contrast with hot carrier injection programming, where a voltage is applied between the source and the drain regions.

[0078] A flash memory array which uses FN tunneling for program and erase is advantageous because thousands of bits in such a flash memory array may be programmed at the same time. Also, FN tunneling is a very efficient way of programming since most (close to 100%) of the current goes to program the device. This is in contrast with hot carrier injection where only about 1-2% of the source-drain current goes to program the device.

[0079] Thus, in an eighth preferred embodiment of the present invention, charge storage devices, such as TFT EEPROMs, are arranged in a flash memory array configuration which utilizes FN tunneling programming. The TFT EEPROMs may be arranged in the rail stack, pillar or self-aligned TFT or configurations of the previous embodiments. Preferably, the TFT EEPROMs are arranged in the rail stack configuration.

[0080] The VGA is not compatible with FN tunneling since the whole channel polysilicon inverts along the length of the pulsed-high word line and will then program cells in addition to the one that needs programming. Therefore, the FN tunneling rail stack (crosspoint) flash array differs from the VGA in that in the FN tunneling array the active polysilicon layer is patterned into polysilicon islands to allow FN tunneling programming. Thus, an extra photolithographic masking step is added to the process of making the rail stack array shown in Figure 5 during which the polysilicon active layer is etched into islands in each device cell. The same photoresist mask can be used to define (i.e., etch) the charge storage regions in each cell.

[0081] In Figure 10A, the flash memory array 330 is preferably formed over a planarized interlayer insulating layer 331, such as a CMP planarized silicon oxide layer. Layer 331 is formed over a substrate (not shown) as in the previous embodiments. Each device of the array (shown by dashed lines 332 in Figure 10A) is thus a TFT because it is formed over an insulating layer.

[0082] The array 330 contains a first plurality of spaced-apart conductive bit lines 333 disposed at a first height above the substrate in a first direction. The array also contains a second plurality of spaced-apart rail-stacks 335. The rail stacks are disposed at a second height in a second direction different from the first direction. Preferably, the bit lines

333 and the rail stacks 335 are arranged perpendicular to each other. The TFT EEPROM 332 is formed at the intersection of the rail stacks 335 and the bit lines 333.

**[0083]** Each rail-stack 335 includes a plurality of semiconductor islands 337, which comprise the active regions of the TFT EEPROMs 332. One surface of the islands 337 is in contact with the bit lines 333. Each rail stack 335 also includes a conductive word line 339 and a charge storage region 341 disposed between a second surface of the semiconductor islands 337 and the word line 339.

**[0084]** The semiconductor islands 337 preferably comprise polysilicon of a first conductivity type (i.e., P- or N-). However, the islands may comprise amorphous silicon if desired. The polysilicon islands 337 include source and drain regions 343 of a second conductivity type (i.e., N+ or P+). The source and drain regions 343 are located at contacting intersections between the bit line conductors 333 and the rail stacks 335.

**[0085]** The bit lines 333 preferably comprise polysilicon of the second conductivity type (i.e., N+ or P+). The bit lines 333 contact the source and drain regions 343. Preferably, the source and drain regions are formed by outdiffusion of dopants from the bit lines. Furthermore, an optional metal or a metal silicide layer (not shown in Figure 10A) may be disposed in contact with the bit lines 333 to increase the conductivity of the bit lines. The space between said spaced-apart bit line conductors 333 is filled with a planarized insulating filler material 345, such as silicon oxide.

**[0086]** The charge storage regions 341 may comprise a dielectric isolated floating gate, electrically isolated nanocrystals or an ONO dielectric stack, as in the previous embodiments. An exemplary array having a dielectric isolated floating gate is illustrated in Figures 10A and

B. Thus, in the example of Figures 10A and B, the charge storage region 341 comprises a polysilicon floating gate 347 between a tunnel dielectric 349, such as a silicon oxide layer, and a control gate dielectric 351 (also known as the intergate or interpoly dielectric) made of a material such as silicon oxide or an ONO layer stack.

**[0087]** As shown in Figures 10A and B, the lateral sides 353 of the tunnel dielectric 349 and the floating gate 347 are aligned to the lateral sides 355 of the semiconductor islands 337. The control gate dielectric 351 extends between the semiconductor islands 337 and contacts the planarized insulating material 345 between the semiconductor islands 337. If desired, the floating gate 347 may be made from hemispherical grain polysilicon which has a textured surface to maximize the control gate to floating gate coupling. Alternatively, the coupling may be increased by increasing the floating gate height, by forming horns or protrusions in the floating gate, or by roughening the floating gate surface.

**[0088]** The word line 339 comprises a polysilicon layer of a second conductivity type (i.e., N+ or P+) and a metal or a metal silicide layer in contact with the polysilicon layer. The word line 339 acts as a control gate of the TFT EEPROM in locations where it overlies the charge storage regions 341. Thus, formation of a separate control gate for each TFT is not required.

**[0089]** In one preferred aspect of this embodiment, the rail stacks 335 are disposed above the bit lines 333, as shown in Figures 10A and B. However, if desired, the rail stacks 335 may be disposed below the bit lines 333 in each device level, (i.e., bottom gate TFT EEPROMs are formed).

[0090] As shown in Figure 10B, the word line 339, the charge storage regions 341 and the semiconductor islands 337 (i.e., the rail stacks 335) are aligned in a plane 356 perpendicular to the substrate and parallel to a source to drain direction. The rail stacks 335 are separated by a second planarized insulating layer 357, such as silicon oxide.

[0091] While the flash memory array may comprise a two dimensional array, preferably, the flash memory array comprises a monolithic three dimensional array comprising a plurality of device levels. For example, three device levels are shown in Figure 10A. The device levels are separated by an interlayer insulating layer 359, such as a silicon oxide layer. If desired, layers 357 and 359 may comprise the same silicon oxide layer which is deposited above and between the rail stacks 359, and then planarized by CMP.

[0092] To program the selected TFT EEPROM 332, either its drain bit line or its source bit line 333 (or both) are grounded while the positive programming voltage is applied to the selected word line 339 adjacent to the device 332 (which is a high impedance node). All other word lines on the same device level are grounded while all other bit lines on the same level device can float or are placed at a slight positive voltage. This means that only the selected cell 332 experiences the programming voltage across it. Through capacitive coupling, the floating gate 347 is pulled high while the source and/or drain 343 are grounded. Electrons tunnel to the floating gate 347 from the source and/or drain 343 and an inversion channel is formed in the silicon channel 337. The current to program such a cell to get a threshold voltage shift of about 5V in approximately one millisecond is several picoamps.

[0093] To erase the cell, the same bit lines 333 can be grounded and a negative voltage pulse is applied to the selected word line 339. All other

word lines can either be grounded or can float. All other bit lines float or are placed at a slight negative voltage. A plurality (or all) of EEPROM cells in the array can be erased at the same time by pulsing a plurality of word lines to a high negative value while all bit lines are grounded. Alternatively, the selected wordline is grounded while the selected cell's bit lines are pulsed positive. All other word lines float or are pulsed slightly positive while all the other bitlines are grounded.

**[0094]** The driver circuit(s) 2 may comprise conventional CMOS thin film transistors formed in an SOI substrate or bulk CMOS transistors formed in a compound semiconductor substrate. Alternatively, Figure 11 illustrates a CMOS TFT array according to a ninth preferred embodiment of the present invention that may be used in a driver circuit 2.

**[0095]** The NMOS and PMOS transistors of the CMOS array may be formed adjacent to each other in the same device level in an alternating fashion (i.e., as alternating NMOS and PMOS transistors). However, in a preferred aspect of the ninth embodiment of the present invention, the one charge carrier type transistors (i.e., NMOS or PMOS) are formed above the other charge carrier type transistors (i.e., PMOS or NMOS) with a common gate line (also known as a word line in memory devices) between them. Thus, the driver circuit 2 of the ninth preferred embodiment comprises a plurality of vertically stacked, common gate CMOS TFT transistors.

**[0096]** One device level of a vertically stacked, common gate CMOS array in a rail stack configuration according to the ninth preferred embodiment of the present invention is shown in Figure 11. It should be noted that the array may also be arranged in a pillar or self-aligned TFT configurations illustrated in Figures 8 and 9. The CMOS array in Figure 11 is similar to the array illustrated in Figure 5, except that transistors of

different charge carrier type are formed on either side of the gate line. In Figure 11, the NMOS transistors are arranged below the PMOS transistors. However, it should be understood that the PMOS transistors may be arranged below the NMOS transistors if desired.

**[0097]** In Figure 11, the array of CMOS devices 400 is preferably formed over a planarized interlayer insulating layer 401, such as a CMP planarized silicon oxide layer. Layer 401 is formed over a substrate (not shown) as in the previous embodiments. Each CMOS device is thus a CMOS TFT because it is formed over an insulating layer.

**[0098]** The array includes a plurality of gate lines (i.e., word lines) 403 (only one gate line is shown in the cross sectional view of Figure 11). Preferably the gate line comprises a first N+ polysilicon layer 405, a silicide layer 407, such as a $TiSi_x$ or $WSi_x$ layer, over the first polysilicon layer and a second P+ polysilicon layer 409 above the silicide layer. The gate line 403 acts as a gate electrode in each TFT. Thus, no separate gate electrodes connected to the gate lines are required.

**[0099]** A first insulating layer 411 is disposed adjacent to a first side of the gate electrode 403. This insulating layer 411 may be a conventional gate dielectric. Preferably, the insulating layer 411 is a charge storage layer (i.e., charge trapping media), such as an ONO stack or isolated nanocrystals, to form charge storage CMOS TFTS, such as EEPROM CMOS TFTs. If floating gate type EEPROM CMOS TFTs are desired, then a floating gate and a control gate dielectric may be added between the insulating layer 411 and the gate line 403.

**[0100]** A p-type semiconductor layer 413, such as a P- polysilicon layer, is disposed on a side of the first insulating layer opposite to the gate 403. This layer contains the NMOS TFT bodies. N+ source and

drain regions 415 are disposed in layer 413. The portions of layer 413 between regions 415 comprise NMOS TFT channel regions.

**[0101]** Preferably, the source and drain regions 415 are formed by outdiffusion of n-type dopants from the source and drain electrodes (i.e., bit lines) 417. However, regions 415 may be formed by any other method, such as by masking and ion implantation. The electrodes 417 contact the source and drain regions 415 and are disposed on the bottom of the p-type semiconductor layer 413 (i.e., on the side of layer 413 opposite to the first insulating layer 411). Preferably, the electrodes 417 comprise N+ polysilicon rails which extend in a direction perpendicular to the gate line 403. If desired, an optional metal or metal silicide layer is formed in contact with electrodes 417 to increase their conductivity. However, the electrodes 417 may comprise metal or metal silicide instead of the heavily doped polysilicon, if desired. A planar insulating filler layer 418, such as silicon oxide, is disposed between the source and drain electrodes 417.

**[0102]** Thus, each NMOS TFT 419 is located between adjacent source and drain regions 415 and comprises a portion of layers 405, 411, 413 and 417, as illustrated in Figure 11. The PMOS TFTS 421 are located above the NMOS TFTs 419.

**[0103]** The PMOS TFTs 421 include a second insulating layer 423 adjacent to a second side of the gate electrode 403. In Figure 11, layer 423 is located on the P+ polysilicon layer 409 of the gate line 403. The insulating layer 423 may be a conventional gate dielectric. Preferably, the insulating layer 423 is a charge storage layer (i.e., charge trapping media), such as an ONO stack or isolated nanocrystals, to form charge storage CMOS TFTS, such as EEPROM CMOS TFTs. If floating gate type EEPROM CMOS TFTs are desired, then a floating gate and a control gate

dielectric may be added between the insulating layer 423 and the gate line 403.

[0104] An n-type semiconductor layer 425, such as an N- polysilicon layer, is disposed above the second insulating layer 423. Layer 425 is disposed on the opposite side of layer 423 from the gate electrode 403. P+ source and drain regions 427 are disposed in layer 425, such that regions of layer 425 between the source and drain regions 427 comprise channel regions of PMOS TFTs. Source and drain electrodes 429 are disposed over the N- polysilicon layer 425 and in contact with the source and drain regions 429. Thus, the electrodes 429 are disposed on top side of the N- polysilicon layer 425 opposite to the second insulating layer 423. A planar insulating filler layer 431, such as silicon oxide, is disposed between the source and drain electrodes 429. If desired, an optional metal or metal silicide layer is formed in contact with electrodes 429 to increase their conductivity.

[0105] Thus, each PMOS TFT 421 is located between adjacent source and drain regions 427 and comprises a portion of layers 409, 423, 425 and 429, as illustrated in Figure 11. A TFT EEPROM CMOS device (419 and 421) is formed at each intersection of the first and the third spaced-apart electrodes or conductors 417, 429 and the common gate line 403. If desired, the CMOS structure may be inverted and the PMOS TFTs formed below NMOS TFTs. It should be noted that NMOS and PMOS electrodes (i.e., bit lines) do not have to fall directly on top of each other, although they preferably should have the same pitch. NMOS and PMOS transistors thus can have different channel lengths, but the pitch (and thus array size) will be limited by the longer of the two channel lengths. In one preferred aspect, TFTs of one conductivity type (i.e., NMOS or PMOS TFTs) contain a charge storage layer or region, while TFTs of the other conductivity type (i.e., PMOS or NMOS) do not have a charge storage

region or layer. Thus, the CMOS of this aspect comprises one EEPROM TFT and one non-EEPROM TFT.

[0106] The TFT CMOS device array 400 illustrated in Figure 11 is highly planar and compact. The NMOS source and drain electrodes 417 comprise polysilicon rails which extend above the interlayer insulating layer 401 in a first plane parallel to the substrate surface. The p-type polysilicon layer 413 extends above the source and drain electrodes 417 in a second plane. The gate line 403 extends above layers 417, 413 and 411 in a third plane. The n-type polysilicon layer 425 extends above the gate line 403 in a fourth plane. The PMOS source and drain electrodes 429 comprise polysilicon rails which extend above the n-type semiconductor layer 425 in a fifth plane. Each of the five planes does not intersect any of the other planes.

[0107] The TFT CMOS array 400 is also self aligned. The gate electrode 403, the first insulating layer 411, the p-type semiconductor layer 413, the second insulating layer 423 and the n-type semiconductor layer 425 comprise a rail stack which is located in a plane parallel to the substrate. The rail stack extends perpendicular to the source and drain electrodes 417, 429. Thus, the gate electrode 403, the first insulating layer 411, the p-type semiconductor layer 413, the second insulating layer 423 and the n-type semiconductor layer 425 are self aligned in a plane perpendicular to the substrate and parallel to the source to drain direction.

[0108] As shown in Figure 12, the resulting TFT CMOS array is a matrix of NMOS 419 and PMOS 421 devices with common gates 403. The array shown in Figure 12 is an unprogrammed or unconfigured array. The array can then be configured into logic elements or memory devices by rupturing the gate dielectric (i.e., the charge storage film or region) to

form a conductive link which connects the gate lines (i.e., word line rows) 403 and source and drain electrodes 417, 429 (i.e., bit lines), or by storing charge in the charge storage regions of either NMOS or PMOS transistors to raise their threshold voltages and keep them permanently off. The array of TFT CMOS EEPROM devices 400 may be used to form either logic elements or a memory array. Furthermore, the same semiconductor device in the unconfigured array may be used either as an antifuse or as an EPROM or an EEPROM.

**[0109]** According to an aspect of the ninth preferred embodiment of the present invention, a driver circuit comprising a plurality of charge storage devices and a plurality of antifuse devices is provided. The circuit may comprise a field programmable gate array or a programmable logic device. Preferably, the plurality of charge storage devices and the plurality of antifuse devices comprise a same set of devices. This greatly simplifies the fabrication of the circuit. These devices function as charge storage devices when a first programming voltage is applied to the devices to turn these devices off by increasing their threshold voltage. These devices also function as antifuses when a second programming voltage higher than a first voltage is applied to the devices. The second voltage may be any voltage which is sufficient to form a conductive link through the charge storage region. For example, the first (i.e., charge storage voltage) may be less than 5 volts, while the second voltage sufficient to form the conductive link may be 5-50 volts, depending on the device characteristics. However, if desired, charge storage and antifuse semiconductor devices having a different structure may be provided.

**[0110]** It should be noted that any charge storage devices which function as an antifuse when a conductive link has been formed through its charge storage region are within the scope of the ninth preferred

embodiment. Thus, any device is within the scope of the ninth preferred embodiment if the device contains a semiconductor active region, a charge storage region adjacent to the semiconductor active region, a first electrode and second electrodes, and where charge is stored in the charge storage region when a first programming voltage is applied between the first and the second electrodes, and a conductive link is formed through the charge storage region to form a conductive path between the first and the second electrodes. Therefore, a charge storage device which is capable of being used as an antifuse is not limited to rail stack TFT EEPROMs. Such charge storage devices may include the pillar or self aligned TFT EEPROMs and diodes with charge storage regions of the previous embodiments.

**[0111]** Figure 13 illustrates how a 4x4 cell array of the circuit of Figure 12 can be programmed into an inverter 443. First, a high voltage is applied between gate (i.e., word) line 445 and bit lines 447, which will be used to carry the output voltage, $V_{out}$. This causes conductive antifuse links 448 to form to electrically connect lines 445 and 447. Then, a programming voltage is applied to all other transistors 450 to increase their threshold voltage to turn them off, except to NMOS transistors 455 and PMOS transistors 457. The NMOS 455 and PMOS 457 transistors form the inverter. When a high voltage, $V_{in}$, is provided into gate line 449, then a low voltage, $V_{out}$, is read out, and vice-versa. Voltages $V_{SS}$ (i.e., ground) and $V_{DD}$ (i.e., power supply voltage) are provided into bit lines 451 and 453 which are connected to transistors 455 and 457.

**[0112]** Figure 14 illustrates how a 4x4 cell array of the circuit of Figure 12 can be programmed into a two input NAND gate 460. First, a high voltage is applied between gate (i.e., word) line 445 and bit lines 447, which will be used to carry the output voltage, $V_{out}$. This causes conductive antifuse links 448 to form to electrically connect lines 445

and 447. Then, a programming voltage is provided to all other transistors 450 to increase their threshold voltage to turn them off, except for PMOS transistors 461 and 465 and NMOS transistors 463 and 465. The transistors 461, 463, 465 and 467 form the NAND gate. Input voltages $V_{in1}$ and $V_{in2}$ are provided into gate lines 469 and 471. CMOS 461/463 is connected to gate line 469, while transistors 465 and 467 are connected to gate line 471. Voltages $V_{SS}$ and $V_{DD}$ are provided into bit lines 473 and 475. NMOS 467 is connected to bit line 475, while PMOS 461 and 465 are connected to bit line 473. Output voltages can be read out from lines 445 or 447, which are connected by a blown antifuse 448.

**[0113]** By forming the driver circuits in an SOI or a compound semiconductor substrate, numerous advantages may be realized. For example, by forming the driver circuits in SOI substrates, the leakage current between the devices of the driver circuits is improved compared to the devices formed in a monocrystalline silicon substrate. Furthermore, the radiation hardness of the devices formed in an SOI substrate is improved compared to the radiation hardness of the devices formed in a monocrystalline silicon substrate.

**[0114]** By forming the driver circuits in a silicon carbide substrate, the leakage current and radiation hardness of the driver circuits is improved compared to the driver circuits formed in a silicon substrate. Furthermore, the driver circuits formed in a silicon carbide substrate can withstand a higher operating voltage than the driver circuits formed in a silicon substrates. Thus, the memory array of the preferred embodiments may be used in an environment where it will be subjected to high radiation doses, such as in military devices and in space craft, by forming the driver circuits in an SOI or a silicon carbide substrate.

[0115] By forming the driver circuits in III-V semiconductor substrates, such as GaAs substrates, the operating speed of the driver circuits is higher than the operating speed of the driver circuits formed in a silicon substrate. Thus, the memory array of the preferred embodiments may be used in an environment where high device operating speed is required.

[0116] An additional advantage of forming driver circuits in SOI or compound semiconductor substrates is that the memory array may be monolithically integrated (i.e., formed over the same substrate) with additional devices which are ordinarily formed on SOI or compound semiconductor substrates. For example, as shown in Figure 15, the driver circuit(s) 2, the memory array 4 and the additional device(s) 500 are monolithically formed on the same substrate, which contains an insulating surface 3. Thus, the additional device 500 is formed over the same substrate as the driver circuits 2 by depositing a semiconductor layer over the substrate and forming the device 500 in the deposited semiconductor layer.

[0117] In one preferred aspect, the additional device 500 which is ordinarily formed on SOI or compound semiconductor substrate is a non-memory device, such as an optoelectronic component. An optoelectronic component may comprise a laser, an light emitting diode (LED) or a semiconductor photodetector. These devices are ordinarily formed on a III-V semiconductor, glass, plastic or ceramic substrates. Thus, the driver circuit 2 is also formed on a III-V, glass, plastic or ceramic substrate.

[0118] In another preferred aspect, the additional device 500 which is ordinarily formed on a compound semiconductor substrate is a microwave circuit (such as a monolithic microwave integrated circuit or

MMIC) or a radio frequency circuit. Thus, the driver circuit 2 is also formed in a III-V substrate.

[0119] In another preferred aspect, the additional device 500 which is ordinarily formed on an SOI substrate, such as a glass, plastic or ceramic substrate, is a liquid crystal display ("LCD"). Since an LCD usually has TFT driver circuits and TFTs in the display matrix, it is preferable to form the driver circuits 2 of the memory array during the same step as the LCD driver circuits, and to form the TFT EEPROMs of one level of the seventh embodiment during the same step as the matrix TFTs of LCD.

[0120] In another preferred aspect, the additional device 500 which is ordinarily formed on an SOI substrate, such as a flexible plastic substrate is a smart card processing circuit. Thus, the memory array and the smart card chip may be formed on the same plastic substrate for use as a smart card.

[0121] In the various embodiments described above, a metal silicide layer was formed in contact with a silicon layer, such as a polysilicon word line or bit line. One preferred method of forming a titanium silicide layer in contact with a silicon layer is by using a silicon cap and a TiN layer. The titanium silicide layer is formed on an undoped amorphous silicon cap layer. The cap layer is formed on a heavily doped silicon layer, such as a polysilicon or amorphous silicon layer doped to a concentration in excess of $10^{19}$ $cm^{-3}$, such as $10^{19}$ $cm^{-3}$ to $10^{21}$ $cm^{-3}$. The cap layer is preferably deposited on P+ polysilicon or N+ amorphous silicon layers. The N+ amorphous silicon may then be recrystallized into N+ polysilicon during subsequent annealing steps.

[0122] A method of forming a titanium silicide ($TiSi_2$) layer comprises the following steps. A heavily doped polysilicon layer is

deposited. For example, a P+ polysilicon layer is boron doped to a concentration of $5x10^{20}$ $cm^{-3}$, and has a thickness of about 1400 Angstroms. A cap layer of undoped amorphous silicon is deposited on the P+ polysilicon layer. The cap may be 600 Angstroms thick, for example. A titanium layer is deposited on the cap. The titanium layer may be 250 Angstroms thick, for example. A titanium nitride layer is deposited on the titanium layer. The titanium nitride layer may be 100 Angstroms thick, for example. Other layer thicknesses may be used, as required.

**[0123]** The layers are annealed at a temperature below 650 °C for less than five minutes to react the titanium and the silicon in the cap to form a C49 phase $TiSi_2$ layer. The anneal may be carried out at 600 °C for 1 minute, for example. If desired, another P+ polysilicon layer is deposited over the stack and the stack is etched into a thin "wire" or "rail", such as a word line or bit line. The wire or rail may be 0.25 mm wide or less. The titanium silicide is then transformed from the C49 to the C54 phase by a high temperature (i.e., above 650 °C) anneal. The anneal can take place before or after the wires or rails are patterned, at 800°C for one minute, for example. By annealing each Si/Ti/TiN film stack below 650°C, dopant diffusion and thermal grooving of the $TiSi_2$ is minimized. Multiple film stacks can be deposited and etched sequentially.

**[0124]** The foregoing description of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention. The drawings and description were chosen in order to explain the principles of the invention and its practical application. The drawings are not necessarily to scale and illustrate the device in schematic block format. It is intended that the

scope of the invention be defined by the claims appended hereto, and their equivalents.